Evaluation of school improvement through an educational effectiveness model

Van der Werf, Greetje; Creemers, Bert; de Jong, Rob; Klaver, Elizabeth

Published in:
Comparative Education Review

DOI:
10.1086/447617

Evaluation of School Improvement through an Educational Effectiveness Model: The Case of Indonesia's PEQIP Project

Author(s): Greetje Van Der Werf, Bert Creemers, Rob De Jong and Elizabeth Klaver

Source: *Comparative Education Review*, Vol. 44, No. 3 (August 2000), pp. 329-355

Published by: The University of Chicago Press on behalf of the Comparative and International Education Society

Stable URL: https://www.jstor.org/stable/10.1086/447617

Evaluation of School Improvement through an Educational Effectiveness Model: The Case of Indonesia's PEQIP Project

GREETJE VAN DER WERF, BERT CREEMERS, ROB DE JONG, AND ELIZABETH KLAVER

Introduction

For several decades, improving the quality of education has been an important issue for the Indonesian government, especially after it succeeded in expanding access to education for children between 7 and 17 years of age. Since 1977, when primary education became compulsory for children between the ages of 7 and 13, the government became more concerned about qualitative issues. In 1978, the first systematic, nationwide improvement program for primary education was introduced, aimed at raising standards of teaching and learning by improving teachers' qualifications. Other improvement projects focused on books, curricula, and the implementation of an educational support system. Despite these efforts some problems have remained, such as poorly trained teachers, inadequate textbooks and materials, and a weak institutional and budgetary framework. These problems may have caused the relatively poor outcomes of the educational system in the early nineties. Retention and dropout rates were still high, and students' enrollment in secondary education remained unsatisfactory.

To improve these outcomes, in 1992 the government initiated the Primary Education Quality Improvement Project (PEQIP). The objective was to introduce policies and mechanisms for improving the overall quality of primary education. In this sense, PEQIP can be considered as a school improvement project aimed at enhancing school effectiveness.

Between September 1995 and September 1997, a school effectiveness study was conducted to establish effects of PEQIP on student achievement and the factors that explain these effects. An integrated school effectiveness model was developed, which took into account PEQIP inputs and intended outputs, school- and classroom-level indicators of effective schools in Western countries, indicators of effective schools in developing countries, and the local context of schools, teachers, and pupils in Indonesia. The model is a multilevel framework in which the educational system is organized hierarchically by pupils, classrooms, and schools. The research design coincides with this framework. Data are measured at individual pupil, classroom, and school levels, as well as at higher levels of the system,

Comparative Education Review, vol. 44, no. 3.

0010-4086/2000/4403-0004$02.00

and are then analyzed according to the multilevel framework. Appropriate controls were created for pupil analysis. This kind of school effectiveness research is unusual in developing countries. Most research in these countries lacks an adequate multilevel design or is mainly limited to education-production models that concentrate on school inputs and resources without linking these to school and teaching practices and without specifying local conditions.[1] The study described here does both. The research questions are as follows: (1) How large are the differences in student achievement between schools? (2) Are between-school differences related to school participation in PEQIP? (3) Which school and classroom factors, indicating PEQIP implementation, are related to between-school differences in student achievement?

Before describing the school effectiveness model used in the study, the PEQIP project is outlined in order clearly to place its inputs and outputs in the model. More information about the project is given in the full report.[2]

The Primary Education Quality Improvement Project (PEQIP)

This project was launched in 1990 when the Indonesian government asked the World Bank for assistance in improving the quality of education. Started in 1992, PEQIP's first implementation phase was completed in October 1997. It is run by the Central Project Management Unit at the Indonesian Ministry of Education and Culture, assisted by three international consultant groups. The project includes involvement at national, provincial, district, and subdistrict levels, within clusters (a group of schools working together), and at school, classroom, and student levels. Activities for PEQIP took place in six provinces: Aceh, Sulawesi Utara, Sumatra Barat, Yogyakarta, Bali, and Nusa Tenggara Timor (NTT), spanning Indonesian regions that had not profited from earlier projects, both relatively rich and poor. In total, 440 schools participated in PEQIP (between 72 and 78 schools per province), which were grouped in 54 clusters (nine clusters per province).

A variety of activities were carried out during the project. Books and materials were developed, school funds were provided, and experiments with new ways of student grouping and teaching methods were stimulated, while teachers, tutors, principals, supervisors, and educational managers participated in upgrading courses. Teachers, principals, and supervisors also participated in appropriate work groups, which were part of the school cluster scheme. A cluster comprised one core school (*Inti*) and six to nine other schools (*Imbas*). Each core school had a teacher activity center where seminars, training activities, and workshops took place. The center was also

[1] B. Fuller and P. Clarke, "Raising School Effects While Ignoring Culture? Local Conditions and the Influence of Classroom Tools, Rules and Pedagogy," *Review of Educational Research* 64, no. 1 (1997): 178–204.

[2] B. Creemers, G. van der Werf, and E. Klaver, *The Effects of PEQIP Indonesia: Report of the Impact and Cost-Analysis Studies* (Jakarta: Ministry of Education and Culture, 1997).

equipped with audiovisual aids, educational materials, and a library. An important objective of the cluster was to encourage the work groups to support professional development. All PEQIP activities were divided into four main components: (1) teacher development, (2) educational management (including community participation), (3) books and learning materials, and (4) evaluation and monitoring. These four components are variables that consistently affect student achievement in developing countries.[3]

The teacher development component focused on training tutors and subject-matter specialists in new strategies for teaching Bahasa Indonesia (the official Indonesian language), mathematics, science, and social studies. After training, tutors and subject-matter specialists were supposed to transfer their newly acquired knowledge and skills to other teachers. Principals, supervisors, and education managers at subdistrict, district, and provincial levels were trained in how to support the implementation of new teaching strategies in the classrooms.

The educational management component was aimed at training supervisors, educational managers, and principals in how to raise the quality of primary education participation. Methods for improving participation included raising money through parents' donations, thereby increasing the educational role of parents and organizing clusters of schools to support the needs of teachers and principals, and monitoring, evaluation and supervision, procurement, and budgeting.

The component on books and learning materials initially included training activities, but when it became clear that teachers did not have enough time and knowledge to produce learning materials, the focus shifted to the development, organization, and use of libraries, newsletter production, and a better distribution system. Student books and teacher guides were provided to schools, as were science kits, reading kits, globes, and other teaching aids for the core subjects. In some areas, sports, art, and music equipment were provided. *Inti* schools were provided with video cameras, televisions, and overhead projectors.

The evaluation and monitoring component consisted of providing training for supervisors and principals. During the training they learned about recognizing the scope of monitoring and evaluation needs, ways of obtaining and using information to satisfy these needs, and skills in finding, evaluating, and using information for planning, as well as how to create and benefit from community participation.

Although community participation was included in the educational management component, it was treated separately by the project and by the consultants. Training focused on school fund-raising, the contributions of services and supplies in meeting school needs, and community programs

[3] H. M. Levin and M. E. Lockheed, eds., *Effective Schools in Developing Countries* (London: Falmer, 1993).

to stimulate compulsory education, school attendance, and evening study. Training was directed at educational managers, supervisors, principals, and community leaders.

Theoretical Background and Research Model

A research model was constructed to study the effects of PEQIP that took the intended outputs, inputs, and other processes into account as well as the context of schools. The study took the framework of educational effectiveness as a point of departure, a framework that is used in Western countries and has also proved useful for educational research in other parts of the world. Numerous projects are currently being conducted in many countries as the annual International Congress of School Effectiveness and School Improvement shows. The *School Effectiveness and School Improvement* journal publishes manuscripts from all parts of the world. The effectiveness framework is used successfully to pinpoint educational problems and solutions through international comparative research in countries like Hong Kong, Taiwan, Australia, the United Kingdom, the United States, Norway, Ireland, and the Netherlands.[4] Developing countries apply the framework, too, more or less successfully for the analysis of the current situation and as a means to pinpoint educational problems.[5]

The framework of educational effectiveness stresses that the results of education should show at the student level. In the case of PEQIP, its contribution to improving the quality of education must be seen in terms of student achievement. That is, PEQIP can only be evaluated as being effective if the pupils in participating schools achieve better results than those in nonparticipating schools. However, because pupil achievement is strongly influenced by background characteristics like socioeconomic status (SES), intelligence, gender, and motivation, they also have to be taken into account.[6]

Sometimes a distinction is made between "general" and "differential" effectiveness. The first concept refers to the achievements of all pupils in a school, the second to the achievements of specific groups of pupils, such as low SES pupils. This distinction is a result of the debate on "excellence" versus "equity" in Western countries because of the large differences in achievement between pupils from different SES backgrounds. In a developing country like Indonesia, however, the differences in achievement and in pupils' backgrounds are much smaller than those between schools, which is

[4] B. P. M. Creemers and D. Reynolds, "Issues and Implications of International Effectiveness Research," *International Journal of Educational Research* 25 (1996): 257–67.

[5] A. R. Riddell, "Assessing Designs for School Effectiveness Research and School Improvement in Developing Countries," *Comparative Education Review* 41, no. 2 (May 1997): 178–204; M. E. Lockheed and A. M. Verspoor, *Improving Primary Education in Developing Countries* (Oxford: Oxford University Press, 1991).

[6] J. B. Carroll, "A Model of School Learning," *Teachers College Record* 64 (1963): 723–33; H. J. Walberg, "Improving the Productivity of America's Schools," *Educational Leadership* 41, no. 8 (1984): 19–27.

the reason PEQIP aims to improve the achievement level of all pupils in the PEQIP schools. Accordingly, this study is limited to the concept of "general" effectiveness.

A second feature of the educational effectiveness framework is the multilevel approach, where distinctions are made between the different levels of education, such as that among the student level, the classroom level, and the school. In the Indonesian context, other levels such as cluster, subdistrict, district, and province are also important. Earlier studies, which examined factors that explain differences in effectiveness between schools, included elements at the school level, such as educational leadership and an orderly and safe school climate.[7] Later on, however, it was acknowledged that factors at the classroom level were more important for achievement, especially time for learning (the time offered to students to engage in learning) and opportunity to learn (the amount of learning content that is taught within a given time period).[8] Moreover, the long history of research into effective teaching has shown that important factors that stimulate student learning include management of the classroom, expectations of student performance, teacher objectives, structuring of lessons, questioning behavior, and immediate exercise after presentation, as well as evaluation, feedback, and corrective instruction.[9]

Recent educational effectiveness models have integrated the factors at school and classroom levels into multilevel models that provide a theoretical basis and might also enhance the explanatory power of research into educational effectiveness.[10] The core of the models is the teaching and learning process at the classroom level. The main factors at this level reflect those that have appeared to be effective in research among classroom teachers. The models reflect those mentioned by R. R. Edmonds, such as educational leadership, high expectations of student achievement, an emphasis on basic skills, a safe and orderly climate, and frequent assessments of students' progress.[11] They assume that factors beyond the classroom level can have direct effects on student achievement, as well as indirect effects mediated by the teaching and learning process in classrooms. The models also stress that factors outside schools are important for student achievement, as was demonstrated by research into school districts in the United States.[12] Finally, the

[7] R. R. Edmonds, "Effective Schools for the Urban Poor," *Educational Leadership* 37, no. 1 (1979): 15–27.

[8] Carroll; T. N. Postlethwaite and K. N. Ross, *Effective Schools in Reading: Implications for Educational Planners* (The Hague: International Studies in Education Achievement, 1992).

[9] B. P. M. Creemers, *The Effective Classroom* (London: Cassell, 1994).

[10] Ibid.; J. Scheerens, *Effective Schooling: Research, Theory and Practice* (London: Cassell, 1992); S. C. Stringfield and R. E. Slavin, "A Hierarchical Longitudinal Model for Elementary School Effects," in *Evaluation of Educational Effectiveness,* ed. B. P. M. Creemers and G. J. Reezigt (Groningen: Interuniversity Center for Educational Evaluation and Educational Effectiveness, 1992), pp. 35–69.

[11] Edmonds.

[12] J. Crispeels, *Purposeful Restructuring: Creating a Culture for Learning and Achievement in Elementary Schools* (London: Falmer, 1992).

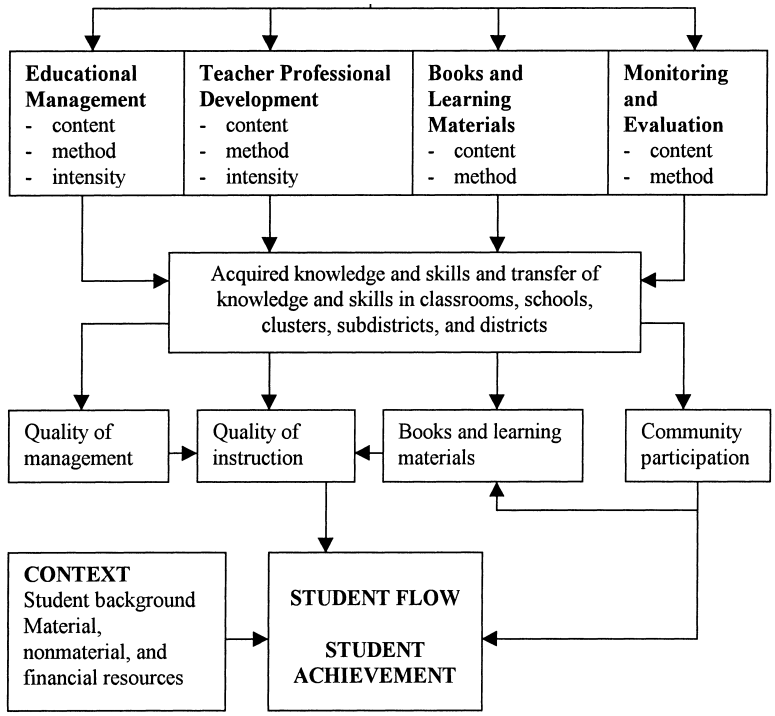

FIG. 1.—Research model

models assume that no one factor or one level is solely responsible for effects at the student level but that the consistency between factors within and across levels contributes to learning outcomes.

In developing a model for the study, we took the integrated model of educational effectiveness developed by B. P. M. Creemers as a point of departure since it elaborates more on the classroom level and is most clearly based on theoretical notions about learning that are supported by the results of empirical research.[13] The model includes factors at different levels that explain student achievement in terms of quality of instruction, time for learning, and opportunity to learn. Within these key concepts, a variety of factors are discerned at each level. The model has variables relating to the context of schools in Indonesia and to the four main PEQIP inputs. It specifies the relationships between the core variables and includes context (material, nonmaterial, and financial resources), inputs (four main PEQIP components), process variables referring to the implementation of PEQIP (quality of management, instruction, books and learning materials, and community participation), and student achievement (fig. 1). The factors at

[13] Creemers.

different levels of the educational system are not expected to influence outcomes directly; rather, they are supposed to provide conditions for effectiveness at the school and the classroom levels. The same holds true for the PEQIP interventions, which are intended to influence outcomes by mediating other factors. The interventions have to be implemented at the cluster and school levels to reach the classroom level. The model takes the PEQIP inputs as a point of departure and arrives at student outcomes via mediating factors.

Core Variables in the Model

The core variables in the model are to be discussed from the top (PEQIP interventions) down to the bottom (output variables) of figure 1.

Content, Method, and Intensity of PEQIP Interventions

Earlier in this article a short summary of each PEQIP input was given. The main intervention in each of these inputs is training. Tutors and subject-matter specialists are trained, and they in turn train classroom teachers. Although the training did not explicitly focus on effective teaching, many aspects reflected educational effectiveness, such as presentation of lessons, advanced organizers, improvement of time on task, monitoring and feedback, and use of a variety of teaching methods. Management training focused on topics such as cluster management, management of resources, community participation, school management, monitoring, evaluation, supervision, procurement, and budgeting. Training regarding books and learning materials offered teachers information on prototype materials, books, and guidelines, the development of prototype teaching aids, and the use of electronic media. All principals were trained in the organization and use of libraries. Evaluation and monitoring training consisted of developing an observation tool with supervisors and providing information on how to use this for improving standards at the classroom and school levels. Training about community participation focused on fund-raising, the contribution of services and supplies to meet current school needs, and community programs to stimulate compulsory education, school attendance, and evening study.

The success of all kinds of training depends on the quality of the training. Indicators to measure the quality of the interventions include content of training (aims of the training), methods (training methods and procedures), and intensity (amount of training received, duration of sessions, frequency of meetings, and guidance).

Acquired Knowledge, Skills, and Transfer to Others

The research model presumes that the PEQIP interventions have differing content, methods, and degrees of intensity, which may lead to differences in the acquired knowledge and skills of the persons trained. The im-

pact of PEQIP also depends on the extent to which teachers, principals, tutors, subject-matter specialists, supervisors, and education managers apply what they have learned. Not only are the actual acquired knowledge and skills important but so is the transfer of these to classrooms, schools, clusters, subdistricts, districts, and provinces. As a consequence of differences in interventions and their transfer, differences may arise regarding the quality of management, instruction, and books and materials, as well as community participation. Indicators to measure this core variable include participant response to training and meetings, how they used the knowledge and skills acquired in their daily practice, and the extent to which they transferred these to others.

Quality of Management

Efficient management at subdistrict, district, provincial, and national levels is important for improving education. Indicators include a policy focusing on effectiveness, an indicator system for evaluation, time-schedule guidelines and supervision, and curriculum guidelines. At the school level, principals can support processes at the classroom level by communicating rules and agreements on classroom instruction, developing an evaluation and monitoring system, supervising and supporting teachers, and communicating the mission of the school to teachers, students and parents. Good principals organize and operate the school timetable efficiently and stimulate teachers to create an orderly and quiet atmosphere and optimize time used in the classrooms.[14]

At the classroom level, teachers demonstrate management and instructional behavior. Management here refers to the activities necessary to make teaching possible. A. Suryadi found that managerial variables that enhance student achievement include attendance at meetings, interaction with parents, classroom supervision by principals, support of principals in drawing up lessons, and regular discussions with supervisors.[15] Quality of management is indicated by the organization of school, cluster, subdistrict, district, and province and the use of monitoring and evaluation systems.

Quality of Instruction

Quality of instruction at the classroom level was included in the Carroll model and elaborated in later models.[16] Quality of instruction influences time for learning and opportunity to learn and is based on the curriculum,

[14] Ibid.

[15] A. Suryadi, *Improving the Educational Quality of Primary Education* (Jakarta: Office of Education and Culture Research and Development, 1992).

[16] Carroll (n. 6 above); R. Glaser, "Components of a Psychology of Instruction: Toward a Science of Design," *Review of Educational Research* 46 (1976): 1–24; S. N. Bennett, "Recent Research on Teaching: A Dream, a Belief, and a Model," *British Journal of Educational Psychology* 28 (1978): 127–47; B. S. Bloom, *Human Characteristics and School Learning* (New York: McGraw-Hill, 1976); J. S. Bruner, *Towards a Theory of Instruction* (New York: Norton, 1966); A. Harnischfeger and D. E. Wiley, "The Teaching Process in Elementary Schools: A Synoptic View," *Curriculum Inquiry* 6 (1976): 5–43.

grouping procedures, and teacher behavior.[17] Indonesian education has a strong tradition of whole-class instruction. However, instruction geared to individual students does not necessarily imply grouping procedures but can also be carried out within whole-class instruction.[18] Some aspects of whole-class instruction affect student outcomes negatively compared to such teaching approaches as textbook reading, use of materials, classroom demonstrations, discussion, and problem-solving methods.[19] Creemers summarized the essential characteristics of teacher behavior as being the structuring of content, clarity of presentation, questioning, immediate exercise after presentation, evaluating whether goals are achieved, and corrective instruction.[20] Elements of quality at the school level include regulations and activities in relation to learning time, regulations concerning training, evaluation and guidance of teachers, regulations concerning student behavior, and regulations regarding the implementation of the curriculum.

Books and Learning Materials

Researchers in Western countries have found that providing more educational equipment does not always contribute to higher student achievement, although it is considered to be an important factor in countries where many schools lack books and learning materials. The study by M. C. S. Moegiadi and W. Elley surprisingly showed that children with insufficient textbooks perform nearly as well as children with books, and sometimes slightly better.[21] Teachers using textbooks did not consistently show better results. Suryadi, however, found that books and materials do contribute to achievement.[22] This might be due to the fact that books and learning materials had improved between when Moediadi and Elley's study was undertaken and when Suryadi's study was done. Indicators in the PEQIP project are the availability and newness of books, the equipment supplied by PEQIP and its use, and the amount and use of books in the library.

Community Participation

Parental and community participation refers to assisting children with homework and engaging in learning activities.[23] In a broader sense it also refers to the mutual collaboration, support, and participation of families, community members, agencies, and school staff in activities and efforts that

[17] Creemers.

[18] P. Croll, "Teaching Methods and Time on Task in Junior Classrooms," *Educational Researcher* 30, no. 2 (1988): 90–97; M. Galton, B. Simon, and P. Croll, *Inside the Primary Classroom* (London: Routledge & Kegan Paul, 1980).

[19] Suryadi.

[20] Creemers (n. 9 above).

[21] M. C. S. Moegiadi and W. Elley, *National Assessment of Primary School Achievement* (Jakarta: Ministry of Education and Culture, 1976).

[22] Suryadi.

[23] O. C. Moles, *Building Home-School Partnerships for Learning* (Washington, D.C.: U.S. Department of Educational Research and Improvement, 1992).

directly and positively affect student outcomes.[24] School-initiated projects that involve parents and students in learning activities at home can enhance attitudes and achievement, or as Chrispeels puts it, "Well-designed parental involvement initiatives, linked to particular curricular areas, can both promote equal educational opportunities and increase school effectiveness."[25] Home and community variables in Suryadi's study affected achievement more strongly than school quality variables, whereas Moegiadi and Elley found the opposite.[26] This difference may be caused by the economic progress that Indonesia has made over the years. In low-income countries, schools affect student achievement more strongly than in middle-income and high-income countries, where the socioeconomic status of the family is an important predictor of achievement. Indicators of community participation in this study include the intensity of contact between schools and parents, contributions of parents to schools, and the engagement of parents in the learning activities of their children.

Student Achievement

In line with the teacher development program, student achievement in mathematics, Bahasa Indonesia, and science is measured in grade 6.

Student Background

To make a fair comparison between PEQIP and non-PEQIP schools, background characteristics of students that are known to affect achievement had to be taken into account. The main characteristics are prior knowledge, intelligence, socioeconomic status (SES), gender, kindergarten attendance, home language, and distance between home and school.

Material, Nonmaterial, and Financial Resources

Literature on educational effectiveness shows disagreement about the importance of resources. According to E. A. Hanushek, material and financial resources do not contribute significantly to outcomes.[27] His study is criticized by L. V. Hedges et al., who used the same data to indicate the importance of resources.[28] A World Bank study in India could not establish the significance of resources.[29] However at a certain stage of development of an educational system, it can be expected that resources will provide conditions

[24] J. Crispeels, "Effective Schools and Home-School Community Partnership Roles: A Framework for Parent Involvement," *School Effectiveness and School Improvement* 7, no. 4 (1996): 297–323.

[25] D. U. Levine and L. W. Lezotte, *Unusually Effective Schools: A Review and Analysis of Research and Practice* (Madison, Wis.: National Center for Effective Schools Research and Development, 1990); Crispeels, "Effective Schools and Home-School Community Partnership Roles."

[26] Suryadi (n. 15 above); Moediadi and Elley.

[27] E. A. Hanushek, "The Economics of Schooling: Production and Efficiency in Public Schools," *Journal of Economic Literature* 24 (1986): 1141–77.

[28] L. V. Hedges, R. D. Laine, and R. Greenwold, "Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes," *Educational Researcher* 23 (1994): 5–14.

[29] World Bank, *Primary Education in India* (Washington, D.C.: World Bank, 1997).

for better education.[30] The Primary Education Quality Improvement Project explored this possibility by financing *Inti* schools more than others. Also, nonmaterial resources, such as the level of education of teachers, may influence achievement, as well as context variables, such as the location or the status of the school or the percentage of low SES pupils.[31]

Method: Sample, Instruments, and Analyses of Data

Sample

The study was conducted in Aceh and Sulawesi Utara, provinces that were selected by the project director of PEQIP at the Ministry of Education and Culture. Within each province, 27 PEQIP schools were selected from all three participating subdistricts. In each of the subdistricts, three clusters of schools participated, each cluster consisting of one core school (*Inti* school) and six to nine satellite schools (*Imbas* schools). For the study, an *Inti* school as well as three *Imbas* schools in each cluster were selected. Thus nine *Inti* schools and 18 *Imbas* schools in each province participated. At the same time, 14 control schools in each province were selected by PEQIP provincial offices. When the PEQIP schools and the control schools are compared regarding status (public or private) and location (rural, urban, suburban), public schools are overrepresented among *Inti* and control schools, while private schools are overrepresented among *Imbas* schools. Also the division with respect to location is not completely equal: rural schools are slightly underrepresented among *Inti* schools. When analyzing PEQIP effects this has to be taken into account.

Instruments

For the study, standardized tests based on the national curriculum were developed for Bahasa Indonesia, mathematics, and science in grade 6. All tests had a multiple choice format. Each test took about 50 minutes. The reliability of the tests is 0.77 (40 items), 0.84 (46 items), and 0.80 (40 items) for Bahasa Indonesia, mathematics, and science, respectively.

The tests for Bahasa Indonesia and mathematics were applied as pre- and posttests. The pretest was administered at the beginning and the posttest at the end of the school year (1996–97). Only a posttest was given for science because this subject is part of the curriculum only in grade 6.

Indicators for student background characteristics were intelligence, socioeconomic status (SES), home language, gender, distance between home and school, and kindergarten attendance. Intelligence was measured during the pretest with a nonverbal intelligence test developed specifically for the study. The test consisted of 46 items for grade 6. The reliability of the

[30] Moediadi and Elley (n. 21 above).
[31] Suryadi (n. 15 above).

test was 0.85. Socioeconomic status was indicated by the educational level of the father (five levels). Home language and kindergarten attendance were dichotomous variables (speaking Bahasa Indonesia or not and having attended kindergarten or not). Distance between home and school was measured in meters. Information about SES and other background characteristics was obtained from teachers during the pretest period.

For measuring the variables with respect to context, resources, inputs, and implementation, a variety of instruments were developed, such as a scheme to analyze documents, interviews, and observation instruments. Most variables were measured by interviews, with different persons providing information on the same variables. The data were collected by trained Indonesian researchers and school supervisors.

The quality of instruction in classrooms for Bahasa Indonesia, mathematics, and science was observed twice by the Indonesian researchers and once by the supervisors. The observation instrument consisted of a low-inference and a high-inference element. The first element consisted of a coding sheet on which the observer coded what was happening during the lesson, minute by minute. Codes included teacher presentation of content, directions for assignments, testing of students, checking of student work, students working with teacher monitors and supports, students working without contact with the teacher, handling of discipline, and other activities (e.g., administrative routines, nonacademic activities, transitions). Every fifth minute, the observers counted the number of students who were involved in the lesson and the number of students who were off-task. The activities and the number of active students were indicators for classroom time (time available for students and time spent by students). The second, high-inference part assessed the frequency of lessons and activities and consisted of a rating scale of 34 statements on classroom organization, aspects of teaching, and the classroom environment. The observers rated these statements on a five-point scale ranging from "never" to "all the time." The amount of instructional content was measured by asking observers to analyze the lesson plans of the teachers and to write down the number of tasks for which instruction was provided. This information was compared with an analysis of classroom books, in which the researchers checked whether all parts of the curriculum were being offered. Finally the observers described teacher activities.

To establish the reliability and validity of the data, comparisons were made between the data collected by the researchers and that collected by the supervisors. The percentage of agreement was about 90 percent on average. However, the correlations between the scores of supervisors and researchers on the high-inference classroom rating scale were too low. Because the ratings of the supervisors were correlated more to school output variables than those of the researchers, we used only the supervisor ratings in

the eventual analyses. Also, data from different sources were cross validated. The interview data were compared with the observations of the researchers and supervisors, as detailed in the school reports they made after each school visit. Data from interviews with different persons were also compared. In case of discrepancies, observations were preferable to interviews, and interviews with independent persons were preferable to data from school-related persons.

Analyses of the Data

The data were analyzed using the VARCL multilevel program separately for Bahasa Indonesia, mathematics, and science.[32] First, an unconditional model, a model without pupil and school variables, was specified in order to establish the proportions of variance to be explained at pupil and school levels. Subsequently, the pupil variables intelligence and SES were specified in the model (Model 1). For Bahasa Indonesia and mathematics, the pretest score was also included. Speaking Bahasa Indonesia at home, kindergarten attendance, distance from home to school, and gender were not included in the model because these variables were either not related to achievement or were related too strongly to intelligence and SES. In Model 2, the school context variables—location (rural = 1; suburban = 2; urban = 3), percentage of pupils who had attended kindergarten, and percentage of low SES pupils—were added to establish the remaining school-level variance. Status of the school (public or private) was not related to achievement, and for this reason, this variable was not included in the model. The remaining school-level variance in Model 2 indicates the differences between schools in student achievement (research question 1).

To explain the differences between schools by their participation in PEQIP (research question 2) in Model 3, the category of school (*Inti, Imbas,* control school) was entered as two dummy variables (*Inti* vs. control; *Imbas* vs. control). If one or both of the dummy variables significantly explained a part of the between-schools variance in achievement, after taking pupil and school context variables into account, and if the regression coefficients of the two dummy variables were positive, we may conclude that participation in PEQIP was effective.

The question of which school and classroom variables, indicating the implementation of PEQIP inputs, are related to the between-schools differences in student achievement is answered in Model 2. In this model, school and classroom variables were added (the model with student and school context variables and without the dummy *Inti* and *Imbas* variables).

[32] The multilevel program VARCL is software for variance component analysis with nested random effects (maximum likelihood); see M. Longford, *VARCL Manual* (Groningen: ProGAMMA, Institute for Software and Internet Expertise, 1993).

TABLE 1
DESCRIPTION OF STUDENT AND SCHOOL CONTEXT VARIABLES, OVERALL AND PER CATEGORY OF SCHOOLS
(Average Scores and Standard Deviations)

	Experimental Group				Control Group		Overall		
	Inti		*Imbas*						
Variable	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Range
Pretest Bahasa Indonesia	21.6	5.4	19.7	5.9	18.4	5.3	19.9	5.7	4–40
Pretest mathematics	19.5	7.5	16.9	7.7	14.6	5.3	17.0	7.2	5–46
Posttest Bahasa Indonesia	22.8	4.8	21.6	5.4	20.0	4.9	21.5	5.2	4–40
Posttest mathematics	24.5	7.8	22.0	8.7	18.8	7.2	21.8	8.2	5–46
Posttest science	24.1	6.0	22.0	6.9	19.4	5.5	21.9	6.5	4–40
Intelligence	27.4	6.7	25.3	7.6	24.3	7.2	25.6	7.3	5–46
Socioeconomic status	3.5	1.0	3.3	1.1	3.1	.9	3.3	1.0	1–5
% low SES pupils	35.5	28.1	53.8	31.9	51.7	28.1	49.1	30.4	0–100
% kindergarten pupils	81.9	12.9	54.9	37.8	44.1	35.2	57.7	36.3	0–100
Location	1.7	.7	1.5	.5	1.6	.7	1.6	.7	1–3

Results

Differences between Schools in Student Achievement and Schools' Participation in PEQIP

First, the student and school context variables in the multilevel analyses conducted to answer the research questions 1 and 2 are described. Table 1 presents the average scores and standard deviations, overall and per category of schools. The pattern is quite clear. Pupils in *Inti* schools score higher on all subjects than pupils in *Imbas* schools, and the latter, in turn, score higher than pupils in control schools. The pattern of student background characteristics (SES and intelligence) and school context variables (with the exception of location) is almost consistent with that of achievement.

Table 2 indicates the results of the multilevel models, including the unconditional Model 0 without pupil and school context variables, Model 1 with only pupil variables, Model 2 with pupil and school context variables, and Model 3 with the added category of school variables (*Inti* vs. control; *Imbas* vs. control). The data are based on 81 schools and 1,854 pupils. The results in Models 1 and 2 show that the pupil and school context variables explain the largest part of the between-schools variance. Including both dummy variables for category of school in Model 3 hardly reduces the unexplained school-level variance and only improves significantly the fit of the model for science. The unstandardized regression coefficients in Model 3 are in table 3. All pupil variables are significant for all three subjects. The effects of the pretest scores—for Bahasa Indonesia and mathematics—are very significant, but intelligence and SES also have a substantial impact. Adding the school context variables to the model reduces the unexplained variance by 2 percent (for mathematics) to 10 percent (for science). The model fit only improves for Bahasa Indonesia and science (see table 2). For Bahasa Indonesia and science, the percentage of low SES pupils is significant and

TABLE 2
PROPORTIONS OF UNEXPLAINED VARIANCE AT PUPIL AND SCHOOL LEVEL, PER SUBJECT

	Model 0	Model 1	Model 2	Model 3
Mathematics:				
Pupil level (%)	54	40	40	40
School level (%)	46	25	23	23
Total (%)	100	65	63	63
Deviance	11,911.98	11,330.79*	11,324.84	11,323.63
Bahasa Indonesia:				
Pupil level (%)	61	45	45	45
School level (%)	39	26	20	20
Total (%)	100	71	65	65
Deviance	10,462.35	10,326.46*	9,887.61*	9,887.47
Science:				
Pupil level (%)	59	53	53	53
School level (%)	41	27	17	14
Total (%)	100	80	70	67
Deviance	10,972.96	10,773.98*	10,704.02*	10,691.53*

NOTE.—Model 0 = model without school and pupil variables; Model 1 = model with pupil variables intelligence, SES, and pretest score (except for science); Model 2 = model with pupil variables and school context variables (location, percent low SES pupils, and percent pupils with kindergarten); and Model 3 = model with pupil and school context variables and category of school.

*Significant improvement of model fit ($P < .05$).

TABLE 3
REGRESSION COEFFICIENTS AND STANDARD ERRORS OF PUPIL AND SCHOOL CONTEXT VARIABLES AND CATEGORY OF SCHOOL, PER SUBJECT
(Model 3)

	Mathematics	Bahasa Indonesia	Science
Grand mean	5.3	11.8	9.7
Pretest score	.52 (.02)*	.38 (.02)*	N.A.
Intelligence	.16 (.02)*	.05 (.01)*	.22 (.02)*
Socioeconomic status (SES)	.51 (.14)*	.24 (.09)*	.62 (.12)*
% kindergarten	.02 (.01)*	.01 (.01)	.02 (.01)*
Location	−.12 (.68)	−.05 (.41)	−1.40 (.42)*
% low SES pupils	−.01 (.01)	−.02 (.01)*	−.02 (.01)*
Category of school:			
Inti	.91 (1.0)	.18 (.61)	2.10 (.62)*
Imbas	1.20 (1.2)	.26 (.76)	2.30 (.77)*

NOTE.—N.A. = not applicable. Numbers in parentheses are standard errors.
*$P < .05$.

negative. For science, the effect of location is positive (the more urban the area of the school, the higher the achievements) and quite large. Location supplants the effect of the pretest score, which is not available for science. With respect to the category of school, table 3 shows that *Inti* and *Imbas* schools score higher than control schools but that the effects are only significant for science. *Imbas* schools score higher than *Inti* schools for all three subjects, but the effects are not significant, with the exception of science.

Table 4 shows the results of the same analyses separately per province. For clarity, we presented only the results of Model 3 (pupil and school variables and category of school). Table 4 shows that, in Aceh, *Inti* and *Imbas* schools scored higher than control schools on all subjects, except for Bahasa Indonesia, where *Imbas* schools scored lower. However, all differences are not significant. In Sulawesi Utara, *Inti* schools scored lower on Bahasa Indonesia and mathematics and higher on science than did control schools. *Imbas* schools there scored higher than control schools on all subjects. The effects for science are significant, both for *Inti* and *Imbas* schools.

In general, we may conclude that PEQIP participation did indeed have effects on student achievement, after taking prior achievement into account, but that these effects are small, not always significant, and not always present for every subject in every category of school. Furthermore, the percentages of unexplained school-level variance in table 4 show that, after controlling pupil background characteristics, prior achievement, school context variables, and school participation in PEQIP, large differences still exist between schools. The next section analyzes to what extent the implementation of PEQIP inputs contribute in practice to explaining these differences between schools.

Explaining Differences in Student Achievement between Schools by the Implementation of PEQIP Inputs

Table 5 shows the variables indicating the implementation of PEQIP inputs at the school and classroom levels, categorized according to management, evaluation and monitoring, teacher development, books and learning materials, and community participation. Table 5 also shows the average scores and standard deviations for each variable, overall and separately for *Inti, Imbas,* and control schools. At first glance, *Inti* and *Imbas* schools do not consistently outscore control schools. On the contrary, control schools score the highest on several variables. A more clear impression on the implementation of PEQIP can be gained by rank ordering the categories of schools for all variables and computing the average rank order score for each category. Control schools receive the highest average rank order score (1.7); *Inti* schools the lowest (2.1). *Imbas* schools hardly differ from *Inti* schools, with an average score of 2.2. Thus we must conclude that PEQIP is not implemented better in PEQIP schools than in control schools, which explains why the participation in PEQIP in general was not very effective.

TABLE 4

Results of Analyses per Province: Model 3 Regression Coefficients and Standard Errors, Percentage of Unexplained Variance, and Improvement of Model Fit Compared to Model 2

	Sulawesi			Aceh		
	Mathematics	Bahasa Indonesia	Science	Mathematics	Bahasa Indonesia	Science
Grand mean	2.7	12.0	10.0	9.3	12.3	11.3
Pretest score	.44 (.03)*	.39 (.03)*	N.A.	.56 (.03)*	.38 (.02)*	N.A.
Intelligence	.24 (.03)*	.05 (.02)*	.22 (.02)*	.10 (.02)*	.05 (.02)*	.21 (.02)*
SES	.63 (.24)*	.18 (.16)	.37 (.19)	.47 (.17)*	.26 (.12)*	.70 (.15)*
% kindergarten	.03 (.01)*	.01 (.01)	.01 (.01)	.01 (.03)	.02 (.01)*	.04 (.02)*
Location	−.63 (1.1)	−.12 (.96)	−1.60 (.72)	.01 (1.1)	.45 (.52)	−.87 (.68)
% low SES	−.01 (.02)	−.03 (.01)*	−.00 (.01)	−.03 (.03)	−.03 (.01)*	−.04 (.01)*
Inti schools	−.53 (1.1)	−.48 (.91)	1.70 (.69)*	2.70 (1.7)	.59 (.78)	1.8 (1.0)
Imbas schools	1.10 (1.3)	.95 (1.1)	2.40 (.85)*	.79 (2.1)	−.76 (1.0)	1.2 (1.3)
Variance pupil level (%)	48	44	65	36	45	45
Variance school level (%)	14	26	11	25	12	14
Total variance unexplained (%)	62	70	76	61	58	60
Difference in model fit[a]	1.6	1.7	8.5*	3.0	2.5	2.4

Note.—N.A. = not applicable. Numbers in parentheses are standard errors.

[a] Difference compared with model 3, the model with pupil and school context variables.

*$P < .01$.

TABLE 5

AVERAGE SCORES AND STANDARD DEVIATIONS OF THE VARIABLES INDICATING THE IMPLEMENTATION OF PEQIP INPUTS, OVERALL AND PER CATEGORY OF SCHOOLS

	Experimental Group								
	Inti		*Imbas*		Control Group		Overall		
Variables	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Range
Teacher professional development:									
Homework:									
Bahasa Indonesia	4.1	1.4	3.8	1.6	4.0	1.2	4.0	1.4	Minutes/day
Mathematics	4.2	1.5	4.0	1.6	4.3	1.2	4.0	1.5	Minutes/day
Science	3.2	1.8	3.9	1.4	4.2	1.2	3.9	1.5	Minutes/day
Time spent on:									
Bahasa Indonesia	463	144	463	174	480	177	469	166	Minutes/week
Mathematics	450	144	465	182	378	55	463	172	Minutes/week
Science	245	68	267	104	286	102	268	97	
Presentation of content	33	14	37	16	35	10	38	16	% lesson time
Pupils working	55	14	51	15	54	11	50	16	% lesson time
Other activities	11	5	11	8	12	4	11	7	% lesson time
Active learning time	76	24	78	20	87	17	78	22	% of pupils
Quality of instruction	3.5	.42	3.3	.32	3.7	.39	3.2	.48	Scale 1–5
Innovative teaching	2.8	.44	2.5	.36	2.8	.37	2.4	.47	Scale 1–5
Frequency of testing	4.5	.54	4.8	.47	4.6	.70	4.7	.62	Scale 1–5
Use of test results	6.7	.56	6.9	.43	6.5	.71	6.7	.81	Scale 1–7
Pupils' attention	2.9	.46	3.0	.00	3.0	.00	2.9	.29	Scale 1–3
Questioning	1.9	.57	1.7	.57	1.9	.57	1.7	.56	Scale 1–3
Comprehension questions	2.2	.76	2.1	.72	2.4	.97	1.9	.77	Scale 1–3
Monitoring work	2.0	.58	2.1	.69	1.8	.63	2.0	.71	Scale 1–3
Grouping of pupils	1.3	.58	1.4	.69	1.2	.42	1.3	.63	Scale 1–3
Management/evaluation:									
Observations in classes	5.1	4.3	5.7	6.9	6.3	9.0	5.7	7.1	Frequency/year
Evaluation of teachers	2.9	.23	2.9	.36	3.0	.00	2.9	.38	Scale 1–3
Evaluation of school quality	2.9	.23	2.8	.38	3.0	.00	2.9	.33	Scale 1–3
Books and learning materials:									
Availability of:									
Equipment	7.7	2.3	3.3	1.4	6.9	3.0	3.9	2.7	Number of materials
Bahasa Indonesia books	78	53	91	80	96	50	90	65	% of pupils[a]
Mathematics books	82	52	92	80	74	41	90	66	% of pupils[a]
Science books	82	43	121	105	78	56	98	80	% of pupils[a]
Community participation:									
Homework regulations	1.9	.23	1.7	.46	1.9	.32	1.7	.44	Scale 1–3
Voluntary work of parents	3.7	2.1	2.0	2.5	3.0	2.4	2.3	2.3	Scale 1–7
Educational involvement of parents	2.7	.32	2.5	.58	2.8	.23	2.5	.51	Scale 1–3
Help with homework	2.5	.51	2.2	.60	2.6	.52	2.3	.57	Scale 1–3
Signing homework	2.8	.42	2.5	.74	2.9	.32	2.6	.64	Scale 1–3
Financial contribution	9.7	5.9	8.1	14.0	10.1	4.4	7.7	10.4	Rp/person × 1,000[b]

[a] Percentage of pupils who have a complete package of books.

[b] Currency of the Indonesian Rupiah in January 1997: Rp 1,000 = US$0.50.

The question remains as to what degree the variables indicating the implementation of PEQIP contribute to the differences in student achievement between schools, whether they are PEQIP or non-PEQIP schools. To answer this question, additional multilevel analyses were conducted in which implementation variables were added to the model with pupil and school context variables (Model 2). Analyses were performed per subject. The province was included as an additional context variable. Table 6 shows the results, with percentages of unexplained variance at the pupil and school levels, total unexplained variance, and difference in fit of the model in comparison with Model 2 and the unstandardized regression coefficients of the student, school, and classroom variables. Variables that are significant ($P < .05$) are indicated by an asterisk (*). Table 6 shows that, for all three subjects, the percentage of unexplained school-level variance was reduced substantially, between 8 and 11 percent, in comparison to Model 2 (see table 2). The variables indicating the implementation of PEQIP account for almost half of the school-level variance that was unexplained after taking student characteristics and school context variables into account. Also, the fit of the models had improved significantly.

However, a number of variables that contribute significantly to explaining school-level variance have a negative impact, and the effects are different across subjects. For mathematics, the table shows positive effects concerning time spent on this subject, the frequency of asking pupils questions during the lessons, the use of test results for evaluating teachers, and school regulations on helping pupils with homework. However, the time teachers spend on other activities (organization, keeping order) instead of presenting learning content or giving exercises to pupils had a negative effect. Also, teacher experience is related negatively to pupil achievement.

For Bahasa Indonesia, innovative teaching (activating pupils, using a variety of teaching methods, encouraging interaction between pupils, and giving them individual tasks) had positive effects, as did school principals' observing teachers in the classroom and voluntary work in schools by parents. As reported before in relation to mathematics, teacher experience had a negative effect on achievement in Bahasa Indonesia.

Unlike the effect on mathematics, the effect of time spent on science had a negative influence on achievement in this subject, as did time the teacher spent on presentation of content and time pupils were engaged in the subject. However, innovative teaching had a positive effect, as did the availability of student textbooks. Particularly striking is that the effect of using test results for evaluating teachers was negative, while using test results for evaluating the quality of a school had a positive effect. Finally, table 6 shows that Sulawesi Utara province scores significantly lower on mathematics and Bahasa Indonesia than Aceh province, while there is no significant difference for science.

TABLE 6
RESULTS OF MULTILEVEL ANALYSES WITH SCHOOL AND CLASSROOM VARIABLES, PER SUBJECT

	Mathematics	Bahasa Indonesia	Science
Grand mean	21.9	18.6	.60
% unexplained variance:			
Pupil level	40	45	53
School level	12	10	9
Total	52	55	62
Improvement model fit[a]	47.1**	46.0**	43.1**
Regression coefficients:			
Student variables:			
Pretest	.51 (.03)*	.38 (.02)*	N.A.
Intelligence	.16 (.02)*	.05 (.01)*	.22 (.02)*
SES	.50 (.14)*	.26 (.09)*	.63 (.12)*
School context:			
% kindergarten	.04 (.02)*	.02 (.01)	.04 (.01)*
% low SES	.00 (.01)	.03 (.01)*	−.01 (.01)
Location	.68 (.84)	.03 (.01)*	−.62 (.56)
Province (Sulut)	−4.6 (1.4)*	−2.1 (.85)*	−.62 (.88)
Implementation variables:			
Teacher professional development:			
Experience teacher	−.15 (.06)*	−.09 (.03)*	.05 (.04)
Homework	−.28 (.29)	−.07 (.16)	−.02 (.18)
Time spent on subject	.01 (.00)*	−.00 (.00)	−.01 (.00)*
Presentation of content	−.03 (.07)	−.04 (.05)	−.04 (.02)*
Pupils working	−.02 (.06)	−.04 (.05)	−.06 (.02)*
Other activities	−.16 (.08)*	−.08 (.07)	.01 (.05)
Active learning time	.00 (.02)	.00 (.01)	.00 (.01)
Quality of instruction	−1.8 (1.9)	−1.7 (1.3)	−.24 (.13)
Innovative teaching	−.14 (1.8)	1.9 (.98)*	1.9 (.97)*
Frequency of testing	−.48 (.73)	−.44 (.43)	.11 (.56)
Use of test results	.01 (.56)	.31 (.34)	−.51 (.41)
Pupils' attention	.54 (1.7)	.36 (.96)	3.4 (2.5)
Questioning	1.7 (.88)*	.28 (.46)	.37 (.53)
Comprehensive questions	.14 (.62)	−.03 (.33)	.44 (.45)
Monitoring work	.72 (.80)	−.33 (.39)	.06 (.57)
Grouping of pupils	−.46 (.73)	−.20 (.47)	.03 (.34)
Management evaluation:			
Observations in classrooms	.06 (.07)	.12 (.05)*	.09 (.07)
Evaluation of teachers	2.5 (1.2)*	.10 (.85)	−2.0 (1.0)*
Evaluation of school quality	2.3 (1.6)	.25 (.95)	2.8 (1.0)*
Books and learning materials:			
Availability of equipment	.16 (.18)	.06 (.10)	−.09 (.12)
Availability of student books	−.01 (.01)	.00 (.00)	.01 (.00)*
Community participation:			
Homework regulations	−1.6 (1.1)	−.23 (.60)	.16 (.71)
Voluntary work of parents	.34 (.21)	.33 (.12)*	.07 (.15)
Educational involvement of parents	−2.5 (2.9)	1.6 (1.8)	−.42 (2.4)
Help with homework	2.9 (1.4)*	1.3 (.89)	1.3 (1.1)*

NOTE.—N.A. = not applicable. Numbers in parentheses are standard errors.
[a] Difference compared to the deviance of model 2, the model with pupil and school context variables.
*$P < .05$.
**$P < .01$.

Conclusions and Discussion

The first research question was, "How large are the differences in student achievement between schools?" With respect to this question, we may conclude that, overall, PEQIP schools only differ significantly from non-PEQIP schools in science achievement. *Inti* and *Imbas* schools score higher than control schools in this subject after taking into account differences in student background characteristics (intelligence, SES, kindergarten attendance) and school location and composition of the student population. However, there are no pretest scores available for this subject. It is highly likely that the positive effects of the participation in PEQIP disappear if pretest scores are taken into account.

When looking at the data per province separately, it appears that the positive effect of PEQIP participation is only significant in Sulawesi Utara for science. But, again, we have to take into account that no pretest scores are available for this subject. For the other two subjects in this province the effects are negative, although not significantly, for *Inti* schools and positive for *Imbas* schools. In Aceh the effects are positive, although not significantly, for all subjects and for both categories of schools (with the exception of Bahasa Indonesia in *Imbas* schools). The effects are not strong. The PEQIP schools score 2.7 points higher at most on a test of 40 items compared to control schools. This is about half a standard deviation. Also, from a cost-effectiveness viewpoint these results are quite disappointing.[33]

There are three possible explanations for the small effects. First, it might be that the effects of PEQIP were already present before the pretests were applied (in the fifth year of the project). In the study, only the additional effects within the sixth year of PEQIP were measured.

Second, it might be possible that the control schools are not comparable with the PEQIP schools. Although they score lower on the pretest and have a student population with a lower socioeconomic background, the quality of the educational processes at the school and classroom levels is even higher than for PEQIP schools. These processes might be responsible for the fact that the control schools achieve an almost comparable learning gain within one school year as PEQIP schools.

The third explanation is that PEQIP was not always successfully implemented. Schools participating in PEQIP did not differ much from non-PEQIP schools with respect to all indicators involved in implementing PEQIP components, and on average they scored even lower. In Aceh, the local authorities not only supported the PEQIP schools but were strongly involved in the control schools as well. They even provided the same books

[33] B. Creemers and G. van der Werf, "Relationships between Educational Effectiveness and the Costs of Interventions: The Case of PEQIP Indonesia" (paper presented at the International Congress of School Effectiveness and School Improvement conference, Manchester, January 3–6, 1998).

and learning materials to the non-PEQIP schools as those the PEQIP schools received from the project. Our impression is that this was not the case in Sulawesi Utara. There, the local authorities were not strongly involved in what happened in the schools, neither in PEQIP nor in non-PEQIP schools.

The strong involvement of the local authorities in schools in Aceh is probably the reason why student achievement in this province is substantially higher than in Sulawesi Utara (and also in other provinces, according to other output data). From an educational effectiveness viewpoint, this is an interesting finding. It supports the assumption in the multilevel models that factors at higher levels also are important conditions for school effectiveness.[34]

The second research question was, "Are between-school differences related to schools' participation in PEQIP?" With respect to this question, we found that the differences in student achievement between schools, after taking pretest scores, student background characteristics, and school context variables into account, were very large, even larger than is found generally in Western countries. Variables indicating the implementation of the PEQIP components (or indicators of effective educational processes) explain about half of the initial school-level variance. The variables that contribute to the explanation differ across subjects, but we can generally conclude that the quality of instruction at the classroom level is important. The effects of innovative teaching are the most consistent because they were found for both Bahasa Indonesia and science. The variable "asking questions to pupils," which is related to innovative teaching, had a positive effect on math achievement.

The findings clearly support the assumption of educational effectiveness models that classroom-level variables are the most important for improving student achievement and that these variables are relevant not only in Western countries but also in developing ones.[35] It also confirms the assumption that an active approach to teaching, in countries where, traditionally, mainly passive learning takes place, is a successful tool for improving student achievement.[36]

Other variables that contribute to explaining differences in student achievement at the school level are the frequency of observations in classrooms by the principal and evaluation of the quality of teachers and the quality of the school. These variables can be considered as indicators of edu-

[34] Creemers (n. 9 above); Scheerens (n. 10 above); M. C. Tsang and C. Wheeler, "Local Initiatives and Their Implications for a Multi-Level Approach to School Improvement in Thailand," in Levin and Lockheed, eds. (n. 3 above), pp. 108–30.

[35] Creemers; Stringfield and Slavin (n. 10 above).

[36] M. E. Lockheed, "The Condition of Primary Education in Developing Countries," in Levin and Lockheed, eds., pp. 20–41.; M. E. Lockheed and H. M. Levin, "Creating Effective Schools," in Levin and Lockheed, eds., pp. 1–20.

cational leadership. Apparently, the activities of principals support good teaching practices at classroom level, which again confirms the assumption in educational effectiveness models that school-level variables are conditional for effective teaching in classrooms.

In addition, the amount of voluntary work undertaken by parents and their engagement in their children's learning activities have positive effects on student achievement. These variables can be considered as a contribution of local resources to the school and as an indicator of parent responsibilities to support and reinforce their children's education. These findings support the idea that community participation is an important concept in theoretical models of educational effectiveness in developing countries.[37]

Only one small effect of resources, books and learning materials, was found in this study. These results are in agreement with those of other studies in Western and developing countries.[38] There are several explanations for the absence of effects of student books and teacher manuals. First, there is not much variation in the availability of books and manuals among schools in Indonesia. Second, any variation that does exist strongly relates to location and composition of the student population of a school. Taking these context variables into account, the effects of books and manuals disappear. Third, it is not the availability of books and manuals that matter but their quality and the way they are used in the teaching.[39] Just like in other developing countries, it was observed that the frequency of using books in classrooms was very low. Apparently, the provision of materials does not guarantee their use.[40]

Finally, the negative effect of teacher experience was found twice in the study. This may be explained by the level of teacher education in Indonesia, which is strongly related to age and experience. Older teachers generally have a lower level of education because they were educated before specific teacher qualifications were required. As may be expected, more qualified teachers have more subject-matter knowledge. Research in developing countries has frequently proved that teachers' subject-matter knowledge has strong positive effects on student achievement.[41] So the teacher's subject-matter knowledge is a concept to take into account in theoretical models of educational effectiveness in developing countries.

We may conclude that programs like PEQIP might make sense in im-

[37] Lockheed and Levin.

[38] Hanushek (n. 27 above); Lockheed; R. Harbison and E. Hanushek, *Educational Performance of the Poor: Lessons from Rural Northeast Brazil* (New York: Oxford University Press, 1992).

[39] Riddell (n. 5 above).

[40] B. Fuller and C. Snyder, "Vocal Teachers, Silent Pupils? Life in Botswana Classrooms," *Comparative Education Review* 35 (1991): 274–94.

[41] B. Fuller, "Raising School Quality in Developing Countries: What Investments Boost Learning?" *Review of Educational Research* 57, no. 3 (1987): 225–92; Suryadi (n. 15 above).

proving the quality of primary education and that improving the teaching process at classroom level seems to be the most promising strategy for realizing higher student achievements.

We also offer certain recommendations for future practice. As was mentioned before, the small effects of PEQIP may be explained by the improper implementation of PEQIP in schools. To improve primary education in Indonesia and in other developing countries in the future, attention has to be paid to achieving a higher degree of implementation of innovations in order to increase the effects. There is still a lot of room for improvement regarding the quality of schools. A lot of school-level variance is unexplained even after taking pupil and school context variables into account. This is in agreement with results from other school effectiveness studies in developing countries.[42]

Practical Implications

Based on the outcomes of the study, together with information from observations in schools and classrooms and with theoretical notions of educational effectiveness, practical implications are suggested for each PEQIP component—that is, teacher development, educational management, evaluation and monitoring, community participation, and books and learning materials. We end with some more general recommendations to improve the quality of education in developing countries.

Educational Management

Management at the school level is most important for improving the quality of education. Management at higher levels must support school-level management. During school observations, many management-related problems appeared, for example, high teacher and student absence, inefficient use of instruction time, and underqualified or unmotivated teachers. The principals of the PEQIP schools focused too much on administrative tasks (keeping records of student results, financial tasks) rather than on educational leadership tasks. This study, as well as international research into educational effectiveness, has shown that effective principals generally undertake the following activities: classroom observation to safeguard the quality of classroom instruction, the creation of conditions for teachers to improve their teaching skills, the selection of high-quality teachers, the guidance of less able ones, the replacement of those who do not improve, the reinforcement of teacher motivation, the control of whether school time is efficiently used, the prevention of teacher and student absence, the control of the implementation of the curriculum in all grades, the promotion of an orderly and friendly school climate, and getting parents and the community in-

[42] Riddell (n. 5 above).

volved effectively in school matters (for details, see "Community Participation," below).

These aspects of effective educational leadership should be included in management training for principals, which means changes in training methods and content. To acquire necessary knowledge and skills, practical exercises are essential so training should include short sessions in small groups. Exercises for principals should be given that can continue to be used in their own schools. In addition, principals need evaluation and feedback on the way these are carried out.

Monitoring and Evaluation

Many principals do not seem to have the knowledge and skills to observe teachers and analyze their performance. They also lack the ability to analyze test results and find it hard to link student progress to the curriculum and quality of instruction. Teaching principals these skills should be part of their management training and can also be recommended for the training of supervisors. Supervisors must also learn how to observe the quality of management at the school level, how to advise on improving principals' management skills, and how to monitor and evaluate schools in terms of student outcomes, standards of instruction, and quality of management.

We recommend separate training for supervisors to provide them with the skills they need. It is unnecessary to train supervisors in community participation and budgeting since these topics are outside their daily practice. Training should be conducted in small groups, last for at least five days, and provide exercises for daily practice. There should be one or two one-day follow-up sessions for evaluation and feedback.

Teacher Development

Observing lessons frequently showed that teachers still practice whole-class instruction. Students are not encouraged to participate actively in lessons. They hardly work individually, with teachers correcting their results later. What teachers learned while training sometimes seemed too remote from their daily practice. The aims of the training may have been unrealistic considering actual practice in most classrooms. Therefore, more realistic training aims should be pursued and should focus more on effective teaching. When teachers have learned how to implement these characteristics, they will better succeed in implementing more innovative approaches. This study has shown that effective teachers use their time efficiently (start lessons on time, keep materials ready, make fast transitions), assign appropriate tasks to students, control whether and how students carry out tasks, motivate students to finish these, provide immediate corrective feedback, keep students actively involved in lessons, explain subject matter clearly and correctly and have substantial knowledge about school subjects, use textbooks and

learning materials, regularly evaluate student tasks and progress, diagnose learning problems, help students who have learning problems, group students when appropriate and on the basis of clear criteria (ability, prior knowledge, interest), assign and correct homework, and provide feedback on the results.

The current PEQIP model (training tutors and subject-matter specialists who, in turn, train teachers) does not seem to be the most effective form of implementation. Training of teachers directly might be more successful, but the costs are much higher. The transfer problem will be reduced when tutors and subject-matter specialists visit schools more often, observe teachers, give in-service training, and are accepted by teachers as experts. This is only possible when tutors and subject-matter specialists are carefully selected for the training, taking into account their positions in the school, their ages, and their educational levels. In addition, the implementation in classrooms of what was learned during the training should be monitored carefully.

Community Participation

Involvement of the community and parents deserves further attention as education becomes more decentralized and the community is invited to participate in its children's school careers. This study stresses the importance of parental involvement and school-home cooperation. However, it showed that strong parental involvement does not automatically lead to positive effects on student achievement and is sometimes even a feature of poor performing schools. It is especially important that parents in the community support schools to prevent student absence and to motivate students to learn, complete their homework, and use their time efficiently. It is not so much the amount and frequency of involvement that is important but the type of involvement. Parents who want to be effectively involved in education can perform the following activities: help their children with their homework, stimulate them to read books, improve their attitude toward regularly attending school, offer help to the school to obtain financial and material resources, and help to maintain an orderly school climate.

We recommend contracts between schools and the community in which parents agree to support the school, and their children and the school agrees to deliver higher levels of achievements. Part of such a contract could perhaps include schools giving parents advice about and support in raising their children. All these aspects of community participation might be included in the training for parents and community leaders, but they should certainly be part of the management training for principals, since effective community participation is a major management task for them.

Books and Learning Materials

Good textbooks used in the classroom may be important for educational improvement. However, this study did not find support for the importance

of textbooks, despite the fact that other studies have shown that textbooks that support teaching contribute to higher standards of education. This study showed that some schools had a lot of equipment and materials that were not always used. Other schools did not have enough books for their students but still seemed to find ways to manage. While teachers appeared to find their own solutions for lack of books, they might have been much more effective if they had more books and learning materials. Thus, more attention should be paid to the distribution of books. The quality of books was not part of this study, but it is worthwhile conducting a study to find out whether teacher and student textbooks support PEQIP's aims. International effectiveness research has shown that high-quality textbooks share the following criteria: goals are explicit and the content is ordered on the basis of these goals, and the textbooks provide a clear learning content structure, methods, and materials for evaluation and feedback, exercises and tasks, and corrective instruction and materials.

We end with some recommendations regarding more general issues concerning the standards of primary education in Indonesia. Teacher quality is a major issue here. This study and others show that a policy directed at employing and hiring qualified teachers may be important for further improvement of education. Educating better teachers and replacing underqualified or less qualified ones takes time but may be accelerated, even though this strategy has financial consequences for the government and personal consequences for teachers.

The SES of students has a major impact on student achievement and student flow. Socioeconomic status is a "container concept," which refers to material and immaterial resources of families. Educational effectiveness studies have shown that SES, together with the ability of students, accounts for most of the variance in student achievement. For this reason, school effect analyses in this study take between-school differences in SES and intelligence of students into account. Schools with high SES students have additional advantages, however. Such schools are often situated in relatively affluent neighborhoods and often employ better-educated teachers. It is recommended that funding be allocated on the basis of a schools' social status because disadvantaged situations need more funding.